WHAT IS CLAIMED IS:

1. A method for maximum a posteriori (MAP) decoding of an input information sequence based on a first information sequence received through a channel, comprising:

iteratively generating a sequence of one or more decode results starting with an initial decode result; and

outputting one of adjacent decode results as a decode of the input information sequence if the adjacent decode results are within a compare threshold.

2. The method of claim 1, wherein the iteratively generating comprises:

a. generating the initial decode result as a first decode result;

b. generating a second decode result based on the first decode result and a model of the channel;

c. comparing the first and second decode results;

d. replacing the first decode result with the second decode result; and

e. repeating b-d if the first and second decode results are not within the compare threshold.

3. The method of claim 2, wherein the generating a second decode result comprises searching for a second information sequence that maximizes a value of an auxiliary function.

4. The method of claim 3, wherein the auxiliary function is based on the expectation maximization (EM) algorithm.

5. The method of claim 4, wherein the model of the channel is a Hidden Markov Model (HMM) having an initial state probability vector π and probability density matrix (PDM) of $\mathbf{P}(X,Y)$, where $X \in \mathbf{X}$, $Y \in \mathbf{Y}$ and elements of $\mathbf{P}(X,Y)$, $p_{ij}(X,Y) = \Pr(j,X,Y \mid i)$, are conditional probability density functions of an information element X of the second information sequence that corresponds to a received element Y of the first information sequence after the HMM transfers from a state i to a state j, the auxiliary function being expressed as:

$$Q(X_1^T, X_{1,p}^T) = \sum_z \Psi(z, X_{1,p}^T, Y_1^T) \log\left(\Psi(z, X_1^T, Y_1^T)\right),$$ where p is a number of iterations, $\Psi(z,\ X_1^T,\ Y_1^T) = \pi_{i_0} \prod_{t=1}^{T} p_{i_{t-1} i_t}(X_t, Y_t)$, T is a number of information elements in a particular information sequence, z is a HMM state sequence i_0^T, π_{i_0} is the probability of an initial state i_0, X_1^T is the second information sequence, $X_{1,p}^T$ is a second information sequence estimate corresponding to a pth iteration, and Y_1^T is the first information sequence.

6. The method of claim 5, wherein the auxiliary function is expanded to be:

$$Q(X_1^T, X_{1,p}^T) = \sum_{t=1}^{T} \sum_{i=1}^{n} \sum_{j=1}^{n} \gamma_{t,ij}(X_{1,p}^T) \log (p_{ij}(X_t, Y_t)) + C$$

where C does not depend on X_1^T and

$$\gamma_{t,ij}(X_{1,p}^T) = \alpha_i(X_{1,p}^{t-1}, Y_1^{t-1}) p_{ij}(X_{t,p}, Y_t)\, \beta_j(X_{t+1,p}^T, Y_{t+1}^T)$$

where $\alpha_i(X_{1,p}^t, Y_1^T)$ and $\beta_j(X_{t+1,p}^T, Y_{t+1}^T)$ are the elements of forward and backward probability vectors defined as

$$\alpha(X_1^t,\ Y_1^t) = \pi \prod_{i=1}^{t} P(X_i, Y_i), \text{ and } \beta(X_1^T, Y_1^T) = \prod_{j=t}^{T} P(X_j, Y_j)\,\mathbf{1},$$ π is an initial probability vector, **1** is the column vector of ones.

7. The method of claim 6, wherein a source of an encoded sequence is a trellis code modulator (TCM), the TCM receiving a source information sequence I_1^T and outputting X_1^T as an encoded information sequence that is transmitted, the TCM defining $X_t = g_t(S_t, I_t)$ where X_t and I_t are the elements of X_1^T and I_1^T for each time t, respectively, S_t is a state of the TCM at t, and $g_t(.)$ is a function relating X_t, to I_t and S_t, the method comprising:

generating, for iteration p+1, a source information sequence estimate $I_{1,p+1}^T$ that corresponds to a sequence of TCM state transitions that has a longest cumulative distance $L(S_{t-1})$ at $t = 1$ or $L(S_0)$, wherein a distance for each of the TCM state transitions is defined by $L(S_t) = L(S_{t+1}) + m(\hat{I}_{t+1}(S_{t+1}))$for the TCM state transitions at each t for

$t = 1, \ldots, T$ and the cumulative distance is the sum of $m(\hat{I}_t(S_t))$ for all t, $m(\hat{I}_t(S_t))$ being defined as

$$m\left(\hat{I}_t(S_t)\right) = \sum_{i=1}^{n_c} \sum_{j=1}^{n_c} \gamma_{t,ij}\left(I_{1,p}^{T}\right) \log p_{c,ij}\left(Y_t \mid X_t(S_t)\right), \text{ for each } t = 1, 2, \ldots, T, \text{ where}$$

$X_t(S_t) = g_t(S_t, \hat{I}_t(S_t))$, n_c is a number of states in an HMM of the channel and $p_{c,ij}(Y_t \mid X_t(S_t))$ are channel conditional probability density functions of Y_t when $X_t(S_t)$ is transmitted by the TCM, $I_{1,p+1}^{T}$ being set to a sequence of $\hat{I}_t$ for all t.

8. The method of claim 7, wherein for each $t = 1, 2, \ldots, T$, the method comprises:

generating $m(\hat{I}_t(S_t))$ for each possible state transition of the TCM;

selecting state trajectories that correspond to largest $L(S_t) = L(S_{t+1}) + m(\hat{I}_{t+1}(S_{t+1}))$ for each state as survivor state trajectories; and

selecting $\hat{I}_t(S_t)$s that correspond to the selected state trajectories as $I_{t,p+1}(S_t)$.

9. The method of claim 8, further comprising:

a. assigning $L(S_T)=0$ for all states at $t = T$;

b. generating $m(\hat{I}_t(S_t))$ for all state transitions between states S_t and all possible states S_{t+1};

c. selecting state transitions between the states S_t and S_{t+1} that have a largest $L(S_t) = L(S_{t+1}) + m(\hat{I}_{t+1}(S_{t+1}))$ and $\hat{I}_{t+1}(S_{t+1})$ that correspond to the selected state transitions;

d. updating the survivor state trajectories at states S_t by adding the selected state transitions to the corresponding survivor state trajectories at state S_{t+1};

e. decrementing t by 1;

f. repeating b-e until $t = 0$; and

g. selecting all the $\hat{I}_t(S_t)$ that correspond to a survivor state trajectory that corresponding to a largest $L(S_t)$ at $t = 0$ as $I_{1,p+1}^{T}$.

10. The method of claim 6, wherein the channel is modeled as $\mathbf{P}_c(Y \mid X) = \mathbf{P}_c\mathbf{B}_c(Y \mid X)$ where $\mathbf{P}_c$ is a channel state transition probability matrix and $\mathbf{B}_c(Y \mid X)$ is a

diagonal matrix of state output probabilities, the method comprising for each $t = 1, 2, \ldots, T$:

generating $\gamma_{t,i}(I_{1,p}^{T}) = \alpha_i(Y_1^t \mid I_{1,p}^t)\beta_i(Y_{t+1}^T \mid I_{t+1,p}^T)$;

selecting an $\hat{I}_t(S_t)$ that maximizes $L(S_t)=L(S_{t+1})+ m(\hat{I}_{t+1}(S_{t+1}))$, where $m(\hat{I}_t(S_t))$ is defined as

$m(\hat{I}_t(S_t)) = \sum_{i=1}^{n_c} \gamma_{t,i}(I_{1,p}^T)\beta_j(Y_t \mid X_t(S_t))$, n_c being a number of states in an HMM of the channel;

selecting state transitions between states S_t and S_{t+1} that corresponds to a largest $L(S_t) = L(S_{t+1}) + m(\hat{I}_{t+1}(S_{t+1}))$; and

forming survivor state trajectories by connecting selected state transitions.

11. The method of claim 10, further comprising:

selecting $\hat{I}_t(S_t)$ that corresponds to a survivor state trajectory at $t = 0$ that has the largest $L(S_t)$ as $I_{1,p+1}^T$ for each pth iteration;

comparing $I_{1,p}^T$ and $I_{1,p+1}^T$; and

outputting $I_{1,p+1}^T$ as the second decode result if $I_{1,p}^T$ and $I_{1,p+1}^T$ are within the compare threshold.

12. A maximum a posteriori (MAP) decoder that decodes a transmitted information sequence using a received information sequence received through a channel, comprising:

a memory; and

a controller coupled to the memory, the controller iteratively generating a sequence of one or more decode results starting with an initial decode result, and outputting one of adjacent decode results as a decode of the input information sequence if the adjacent decode results are within a compare threshold.

13. The decoder of claim 12, wherein the controller:

a. generates the initial decode result as a first decode result;

b. generates a second decode result based on the first decode result and a model of the channel;

c. compares the first and second decode results;

d. replaces the first decode result with the second decode result; and

e. repeats b-d until the first and second decode result are not within the compare threshold.

14. The decoder of claim 13, wherein the controller searches for information sequence that maximizes a value of an auxiliary function.

15. The decoder of claim 14, wherein the auxiliary function is based on expectation maximization (EM).

16. The decoder of claim 15, wherein the model of the channel is a Hidden Markov Model (HMM) having an initial state probability vector π and probability density matrix (PDM) of $\mathbf{P}(X,Y)$, where $X \in \mathbf{X}$, $Y \in \mathbf{Y}$ and elements of $\mathbf{P}(X,Y)$, $p_{ij}(X,Y) = \Pr(j,X,Y \mid i)$, are conditional probability density functions of an information element X of the second information sequence that corresponds to a received element Y of the first information sequence after the HMM transfers from a state i to a state j, the auxiliary function being expressed as:

$$Q(X_1^T, X_{1,p}^T) = \sum_z \Psi(z, X_{1,p}^T, Y_1^T)\log\left(\Psi(z, X_1^T, Y_1^T)\right),$$ where p is a number of iterations, $\Psi(z, X_1^T, Y_1^T) = \pi_{i_0}\prod_{t=1}^{T} p_{i_{t-1}i_t}(X_t, Y_t)$, T is a number of information elements in a particular information sequence, z is a HMM state sequence i_0^T, π_{i_0} is the probability of an initial state i_0, X_1^T is the second information sequence, $X_{1,p}^T$ is a second information sequence estimate corresponding to a pth iteration, and Y_1^T is the first information sequence.

17. The decoder of claim 16, wherein the auxiliary function is expanded to be:

$$Q(X_1^T, X_{1,p}^T) = \sum_{t=1}^{T}\sum_{i=1}^{n}\sum_{j=1}^{n} \gamma_{t,ij}(X_{1,p}^T)\log\left(p_{ij}(X_t, Y_t)\right) + C$$

where C does not depend on X_1^T and

$$\gamma_{t,ij}(X_{1,p}^T) = \alpha_i(X_{1,p}^{t-1}, Y_1^{t-1})\, p_{ij}(X_{t,p}, Y_t)\, \beta_j(X_{t+1,p}^T, Y_{t+1}^T)$$

where $\alpha_i(X_{1,p}^t, Y_1^T)$ and $\beta_j(X_{t+1,p}^T, Y_{t+1}^T)$ are the elements of forward and backward probability vectors defined as

$\alpha(X_1^t,\ Y_1^t) = \pi \prod_{i=1}^{t} P(X_i, Y_i)$, and $\beta(X_{t+1}^T, Y_{t+1}^T) = \prod_{j=t+1}^{T} P(X_j, Y_j)\,\mathbf{1}$, π is an initial probability vector, **1** is the column vector of ones.

18. The decoder of claim 17, wherein a source of an encoded sequence is a trellis code modulator (TCM), the TCM receiving a source information sequence I_1^T and outputting X_1^T as an encoded information sequence that is transmitted, the TCM defining $X_t = g_t(S_t, I_t)$ where X_t and I_t are the elements of X_1^T and I_1^T for each time t, respectively, S_t is a state of the TCM at t, and $g_t(.)$ is a function relating X_t, to I_t and S_t, the controller generates, for iteration p+1, an input information sequence estimate $I_{1,p+1}^T$ that corresponds to a sequence of TCM state transitions that has a longest cumulative distance $L(S_{t-1})$ at t = 1 or $L(S_0)$, wherein a distance for each of the TCM state transitions is defined by $L(S_{t+1}) = L(S_{t+1}) + m(\hat{I}_{t+1}(S_{t+1}))$ for the TCM state transitions at each t for t = 1, ..., T and the cumulative distance is the sum of $m(\hat{I}_t(S_t))$ for all t, $m(\hat{I}_t(S_t))$ being defined as

$$m\left(\hat{I}_t(S_t)\right) = \sum_{i=1}^{n_c} \sum_{j=1}^{n_c} \gamma_{t,ij}\left(I_{1,p}^T\right) \log p_{c,ij}\left(Y_t \mid X_t(S_t)\right),$$ for each t = 1, 2, ..., T, where

$X_t(S_t) = g_t(S_t, \hat{I}_t(S_t))$, n_c is a number of states in an HMM of the channel and $p_{c,ij}(Y_t \mid X_t(S_t))$ are channel conditional probability density functions of Y_t when $X_t(S_t)$ is transmitted by the TCM, $I_{1,p+1}^T$ being set to a sequence of $\hat{I}_t$ for all t.

19. The decoder of claim 18, wherein for each t = 1, 2, ..., T, the controller generating $m(\hat{I}_t(S_t))$ for each possible state transition of the TCM, selecting state trajectories that correspond to largest $L(S_t) = L(S_{t+1}) + m(\hat{I}_{t+1}(S_{t+1}))$ for each state as survivor state trajectories, and selecting $\hat{I}_{t+1}(S_{t+1})$s that correspond to the selected state trajectories as $I_{t+1,p+1}(S_{t+1})$.

20. The decoder of claim 19, wherein the controller:

a. assigns $L(S_T)=0$ for all states at t = T;

b. generates $m(\hat{I}_t(S_t))$ for all state transitions between states S_t and all possible states S_{t+1};

c. selects state transitions between the states S_t and S_{t+1} that have a largest $L(S_t) = L(S_{t+1}) + m(\hat{I}_{t+1}(S_{t+1}))$ and $\hat{I}_{t+1}(S_{t+1})$ that correspond to the selected state transitions;

d. updates the survivor state trajectories at states S_t by adding the selected state transitions to the corresponding survivor state trajectories at state S_{t+1};

e. decrements t by 1;

f. repeats b-e until $t = 0$; and

g. selects all the $\hat{I}_t(S_t)$ that correspond to a survivor state trajectory that corresponding to a largest $L(S_t)$ at $t = 0$ as $I^T_{1,p+1}$.

21. The decoder of claim 20, wherein the channel is modeled as $\mathbf{P}_c(Y \mid X) = \mathbf{P}_c\mathbf{B}_c(Y \mid X)$ where $\mathbf{P}_c$ is a channel state transition probability matrix and $\mathbf{B}_c(Y \mid X)$ is a diagonal matrix of state output probabilities, for each $t = 1, 2, \ldots, T$, the controller:

generates $\gamma_{t,i}(I^T_{1,p}) = \alpha_i(Y^t_1 \mid I^t_{1,p})\beta_i(Y^T_{t+1} \mid I^T_{t+1,p})$;

selects an $\hat{I}_t(S_t)$ that maximizes $L(S_t) = L(S_{t+1}) + m(\hat{I}_{t+1}(S_{t+1}))$, where $m(\hat{I}_t(S_t))$ is defined as

$$m(\hat{I}_t(S_t)) = \sum_{i=1}^{n_c} \gamma_{t,i}(I^T_{1,p})\beta_j(Y_t \mid X_t(S_t)),$$

n_c being a number of states in an HMM of the channel;

selects state transitions between states S_t and S_{t+1} that corresponds to a largest $L(S_t) = L(S_{t+1}) + m(\hat{I}_{t+1}(S_{t+1}))$; and

forms survivor state trajectories by connecting selected state transitions.

22. The decoder of claim 21, wherein the controller selects $\hat{I}_t(S_t)$ that corresponds to a survivor state trajectory at $t = 0$ that has the largest $L(S_t)$ as $I^T_{1,p+1}$ for each pth iteration, compares $I^T_{1,p}$ and $I^T_{1,p+1}$, and outputs $I^T_{1,p+1}$ as the second decode result if $I^T_{1,p}$ and $I^T_{1,p+1}$ are within the compare threshold.